

OLIVER PAJONK, BOJANA V. ROSIĆ, AND  
HERMANN G. MATTHIES

---

# DETERMINISTIC LINEAR BAYESIAN UPDATING OF STATE AND MODEL PARAMETERS FOR A CHAOTIC MODEL

---



INFORMATIKBERICHT 2012-01

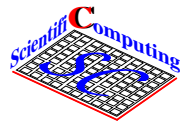
INSTITUTE OF SCIENTIFIC COMPUTING  
CARL-FRIEDRICH-GAUSS-FAKULTÄT  
TECHNISCHE UNIVERSITÄT BRAUNSCHWEIG

Braunschweig, Germany

This document was created in January 2012 using L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>.

Institute of Scientific Computing  
Technische Universität Braunschweig  
Hans-Sommer-Straße 65  
D-38106 Braunschweig, Germany

e-mail: [wire@tu-bs.de](mailto:wire@tu-bs.de)  
url: [www.wire.tu-bs.de](http://www.wire.tu-bs.de)



Copyright © by Oliver Pajonk, Bojana V. Rosić, and Hermann G. Matthies

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted in connection with reviews or scholarly analysis. Permission for use must always be obtained from the copyright holder.

Alle Rechte vorbehalten, auch das des auszugsweisen Nachdrucks, der auszugsweisen oder vollständigen Wiedergabe (Photographie, Mikroskopie), der Speicherung in Datenverarbeitungsanlagen und das der Übersetzung.

# Deterministic Linear Bayesian Updating of State and Model Parameters for a Chaotic Model

Oliver Pajonk, Bojana V. Rosić, and Hermann G. Matthies

Institute of Scientific Computing, TU Braunschweig  
[wire@tu-bs.de](mailto:wire@tu-bs.de)

## Abstract

We present a sampling-free implementation of a linear Bayesian filter. It is based on spectral series expansions of the involved random variables, one such example being Wiener's polynomial chaos. The method is applied to a combined state and parameter estimation problem for a chaotic system, the well-known Lorenz-63 model. We compare it to the ensemble Kalman filter (EnKF), which is essentially a stochastic implementation of the same underlying estimator—a fact which is demonstrated in the paper. The spectral method is found to be more reliable for the same computational load, especially for the variance estimation. This is to be expected due to the fully deterministic implementation.

**Keywords:** Bayesian estimation, polynomial chaos expansion, Kalman filter, inverse problem, white noise analysis, Lorenz-63

**AMS classification:** 60H40, 65M32, 62L12



Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Linear Bayesian Updating</b>	<b>2</b>
2.1	Linear Conditional Expectation . . . . .	2
2.2	Stochastic Implementation: The Ensemble Kalman Filter . . . . .	3
2.3	A Deterministic Implementation: By Polynomial Chaos Expansion .	4
<b>3</b>	<b>Combined Parameter and State Estimation for the Lorenz-63 Model</b>	<b>4</b>
3.1	The Model . . . . .	5
3.2	Experimental Setup . . . . .	6
3.3	Results and Discussion . . . . .	9
<b>4</b>	<b>Conclusion</b>	<b>12</b>
<b>A</b>	<b>The Hermite Algebra</b>	<b>12</b>
<b>B</b>	<b>Higher Order Moments</b>	<b>14</b>
	<b>References</b>	<b>14</b>



# 1 Introduction

Efficient and effective parameter estimation for complex geophysical systems is an important task: all subsequent aspects of an uncertainty quantification and robust optimisation workflow depend on these results. Especially in the common inverse setting—with the available evidence being connected only indirectly to the parameters through noisy measurements of the system state—makes the inference problem harder and requires more sophisticated solution methods.

The two major families of methods for approaching such—usually ill-posed—problems are regularised optimisation methods (cf. [Engl et al. \(1996\)](#)) and Bayesian approaches (cf. [Tarantola \(2005\)](#)), with the topic of this work being part of the latter. A popular Bayesian approach is to sequentially derive estimates for the combined parameter and state estimation problem, where obtaining new estimates of model parameters is integrated into a data assimilation setting for the model state to stabilise the forecast (cf. [Lahoz et al. \(2010\)](#); [Evensen \(2009a\)](#)). In this paper we will examine the capabilities of a recently proposed ([Pajonk et al., 2011](#); [Rosić et al., 2011](#)), fully deterministic linear Bayesian method at estimating parameters of a chaotic, low-dimensional system—the well-known Lorenz-63 model ([Lorenz, 1963](#)).

First approaches to quantify the influence of uncertain parameters in this model include [Fleming \(1993\)](#). There, the authors evaluate the effects using complex moment equations—therefore only the single system parameter  $r$  is uncertain. Also [Lea et al. \(2000\)](#) conduct a sensitivity analysis of Lorenz-63 to perturbations of the parameter  $r$ . Estimation approaches include [Kivman \(2003\)](#), where  $r$  and  $s$  are estimated by the ensemble Kalman filter (EnKF) and another Monte Carlo (MC) method called “sequential importance resampling” (SIR) in a combined parameter-state estimation setting. [Moolenaar & Selten \(2004\)](#) discuss “effective parameter perturbations” for Lorenz-63 by an adjoint method; [Annan & Hargreaves \(2004\)](#) employ an iterative EnKF-based approach with inflation. [Ambadan & Tang \(2009\)](#) perform combined estimation for Lorenz-63 using EnKF and a sigma point approach (but see also the associated discussion in [Hamill et al. \(2009\)](#) and [Tang & Ambadan \(2009\)](#)).

The major drawbacks of MC methods like EnKF and SIR are certainly the slow convergence rate as well as sampling errors arising from the measurement ensemble, which frequently have a significant impact. Additionally, the inherent non-determinism rules out MC methods from certain applications. Approaches towards mitigating the sampling errors of EnKF include the family of ensemble square root filters (e.g. [Tippett et al. \(2003\)](#)), which is based on [Andrews \(1968\)](#). While they avoid the creation of a measurement ensemble, the methods still need the random initial ensemble and can therefore be considered as pseudo-deterministic. [Sakov & Oke \(2008\)](#) is another pseudo-deterministic variant of the EnKF.

The approach discussed in this paper is a *fully* deterministic implementation of

the linear Bayesian estimator (Goldstein & Wooff, 2007) that forms the basis of all Kalman filters—and in fact the original Kalman filter (Kalman, 1960) can be shown to be the low-order part of the presented approach (see Pajonk et al. (2011); Rosić et al. (2011)). Its practical computation is based on orthogonal series expansions of the random variables (cf. Le Maître & Knio (2010); Xiu (2010); Matthies (2007); Matthies et al. (2011) and the references therein). Similar approaches have been discussed previously: Blanchard et al. (2010) is based on extended Kalman filter theory, whereas Saad & Ghanem (2009) is directly derived from Kalman filter theory. Related approaches include Li & Xiu (2009), where a usual EnKF is employed to update coefficients of a specific series expansion; in Pence et al. (2010) such updated coefficients are computed by a maximum likelihood approach.

The linear Bayesian estimation approach, together with the specific implementation employed in this work, is briefly introduced in section 2. There also the fact that the EnKF is a close relative to the presented approach is succinctly illuminated. Section 3 provides a numerical evaluation of this approach in a combined parameter and state estimation setting using Lorenz-63 and compares the results to a usual EnKF implementation. Section 4 concludes the work.

## 2 Linear Bayesian Updating

In Bayesian approaches, like the one considered in this work, limited information is represented by random variables (RVs). These are measurable functions  $\phi, \psi \in L_0(\Omega; \mathbb{R})$  defined on a probability space  $(\Omega, \mathfrak{S}, \mathbb{P})$ . There,  $\Omega$  is an abstract sample space,  $\mathfrak{S}$  a  $\sigma$ -algebra of subsets of  $\Omega$  and  $\mathbb{P}$  a probability measure.

### 2.1 Linear Conditional Expectation

In this setting, the limited information contained in one RV  $\psi(\omega)$  can be related to another RV  $\phi(\omega)$  via the *conditional expectation* (CE)  $\mathbb{E}(\phi | \psi)$ , the expectation of  $\phi(\omega)$  given  $\psi(\omega)$ . It is well known that computing the CE is equivalent to computing conditional probabilities—and in fact, already Kolmogorov (1956) defined conditional probabilities via CE. This CE is the improved estimator one is seeking.

In the special case of RVs with finite variance,  $L_2(\Omega; \mathbb{R})$ , the CE is an orthogonal projection in a Hilbert space (Luenberger, 1969). Additionally, the Doob-Dynkin lemma (Bobrowski, 2005, Eq. 2.1.24) tells us that the CE is a measurable function of  $\psi(\omega)$  (Bobrowski, 2005, p. 90). To make progress we limit ourselves to *linear* (or rather, affine) measurable functions in the  $L_2$ -setting:

$$\mathbb{E}_{lin}(\phi | \psi) = F\psi(\omega) + b. \quad (1)$$



It is by construction the linear estimator with minimum variance, also called the linear least squares approximation. Computing this linear approximation to the conditional expectation (LCE) becomes surprisingly simple and is given by (cf. [Luenberger \(1969, p. 87\)](#), [Whittle \(2000, chapter 14\)](#))

$$\begin{aligned}\mathbb{E}_{lin}(\phi|\psi) &= \underbrace{\mathbf{C}_{\phi,\psi}\mathbf{C}_{\psi}^{-1}}_{=:F} \psi(\omega) + \underbrace{\mathbb{E}(\phi) - \mathbf{C}_{\phi,\psi}\mathbf{C}_{\psi}^{-1}\mathbb{E}(\psi)}_{=:b} \\ &= \mathbb{E}(\phi) + \mathbf{C}_{\phi,\psi}\mathbf{C}_{\psi}^{-1}(\psi - \mathbb{E}(\psi))\end{aligned}\quad (2)$$

where  $\mathbf{C}_{\phi,\psi}$  denotes the co-variance of  $\phi(\omega)$  and  $\psi(\omega)$ , and  $\mathbf{C}_{\psi}$  is a shorthand for  $\mathbf{C}_{\psi,\psi}$ . Comparing this to the full CE, we limit the projection of  $\phi(\omega)$  onto the subspace of *linear* measurable functions of  $\psi(\omega)$ —a subspace of *all* measurable functions.

A generalisation of Eq. (2) of high practical importance is given by the following recursive formula for *updating* an LCE  $\phi(\omega)$  with *additional evidence*  $\psi(\omega)$  (cf. [Luenberger \(1969, p. 93ff\)](#)):

$$\hat{\phi}(\omega) = \phi(\omega) + \underbrace{\mathbf{C}_{\phi,h(\phi)}(\mathbf{C}_{\psi} + \mathbf{C}_{h(\phi)})^{-1}}_{=:K}(\psi(\omega) - h(\phi(\omega))).\quad (3)$$

Here,  $h$  is the so-called *measurement operator* or *forward model* which relates  $\phi(\omega)$  to  $\psi(\omega)$ . In this context,  $\phi(\omega)$  usually describes the state and/or parameters of a numerical model, and  $h(\phi(\omega))$  is the *forecast measurement*. Note that  $K$  is the well-known Kalman gain. For Eq. (3) to be equivalent to a single evaluation of Eq. (2) with all evidence and model quantities concatenated into vectors  $\boldsymbol{\phi}(\omega)$  and  $\boldsymbol{\psi}(\omega)$ , the involved errors have to be uncorrelated (e.g. [Evensen \(2009a, section 7.3.2\)](#)). Eq. (3) is a well-known result and forms the basis of algorithms such as the Kalman filter. There, additional evidence arrives *over time* and the recursive nature of that equation becomes especially convenient.

The main implementation choice which has to be made now is how to actually represent the random variables in practical computations—until now they are only abstract functions.

## 2.2 Stochastic Implementation: The Ensemble Kalman Filter

Arguably the most popular choice to represent RVs in computations is by random (MC) sampling. There, many samples  $\phi(\omega_i)$  are created according to the probability measure  $\mathbb{P}$ . For each of the samples, the forward problem is again deterministic and independent of all other samples. The necessary covariances for computing Eq. (3) can be estimated directly from these samples. This representation leads to the popular ensemble Kalman filter (EnKF) ([Evensen, 2009a,b](#)).

## 2.3 A Deterministic Implementation: By Polynomial Chaos Expansion

Another choice of practical representation of RVs is by orthogonal series expansions. A popular choice are series of polynomials in known, simple RVs, with an example being Wiener's polynomial chaos expansion (PCE) (Wiener, 1938; Holden et al., 2010; Janson, 1997; Malliavin, 1997; Hida et al., 1999; Hida & Si, 2008). There, multivariate Hermite polynomials in standard normal RVs are employed:

$$\phi(\omega) = \sum_{\alpha \in \mathcal{J}} \phi^\alpha H_\alpha(\theta_1(\omega), \dots, \theta_k(\omega), \dots), \quad (4)$$

with  $\mathcal{J} := \mathbb{N}_0^{(\mathbb{N})}$  being a multi-index set discriminating the Hermite polynomials  $H_\alpha$  and coefficients  $\phi^\alpha$ . The sequence of coefficients  $(\phi^\alpha)_{\alpha \in \mathcal{J}}$ —also called the Hermite transform  $\mathcal{H}(\phi)$  of the RV  $\phi$ , see Matthies (2007)—represents the RV and may be computed simply by projection:

$$\forall \alpha \in \mathcal{J} : \quad \phi^\alpha = \mathbb{E}(\phi(\cdot) H_\alpha(\cdot)) / \langle H_\alpha | H_\alpha \rangle. \quad (5)$$

The Cameron-Martin theorem (Malliavin, 1997; Hida et al., 1999) assures us that the polynomial algebra of standard normal RVs is dense in  $L_2(\Omega)$ , i.e. we may write any RV with finite variance as such a series of polynomials.

Given expansions for any involved RV as in Eq. (4), one may efficiently compute moments of these RVs as it is shown in Eq. (21)—note that this includes the covariances necessary for computing Eq. (3) (see Eq. (23)). Therefore, this implementation of Eq. (3) is deterministic in every aspect: no sampling is required at any stage and all quantities can be efficiently computed from the series representation.

For the numerical implementation the Hermite transform obviously has to be limited to a finite number of Gaussian RVs and to a finite polynomial degree, described by a finite index set  $\mathcal{J}_Z \subset \mathcal{J}$ . A simple way is to truncate the series at a certain highest degree of expansion, but a more computationally efficient approach would be to perform an adaptive choice of truncation for  $\mathcal{J}_Z$  (Nouy & Le Maître, 2009; El-Moselhy, 2010; Krosche & Niekamp, 2010). However, this path is not considered further in this paper.

## 3 Combined Parameter and State Estimation for the Lorenz-63 Model

In this section we demonstrate how the PCE-based LCE implementation performs when applied to a low-dimensional, chaotic sequential combined parameter and state estimation problem. We chose the well-known Lorenz-63 model (Lorenz,

1963) as an example, since this has already been studied in similar contexts (Fleming, 1993; Lea et al., 2000; Kivman, 2003; Annan & Hargreaves, 2004; Ambadan & Tang, 2009). The results are compared to a popular Monte-Carlo-based LCE implementation, the EnKF.

### 3.1 The Model

The state evolution of the Lorenz-63 model,  $\dot{u} = \frac{du}{dt} = f(u)$ ;  $u(0) = u_0$ , is described by the following set of ordinary differential equations (ODEs):

$$\begin{aligned}\frac{dx}{dt} &= s(y - x) \\ \frac{dy}{dt} &= rx - y - xz \\ \frac{dz}{dt} &= xy - bz,\end{aligned}\tag{6}$$

with three parameters  $s, r$  and  $b$ . Here we are mainly interested in the dependence of the state on uncertainties in these parameters. Therefore we model them as independent Gaussian RVs:

$$\begin{aligned}s(\omega) &\sim \mathcal{N}(s_0, \sigma_1) \\ r(\omega) &\sim \mathcal{N}(r_0, \sigma_2) \\ b(\omega) &\sim \mathcal{N}(b_0, \sigma_3).\end{aligned}\tag{7}$$

Due to the appearance of RVs, the deterministic model Eq. (6) turns into a system of stochastic differential equations (SDEs, e.g. Øksendal (2003)),

$$\begin{aligned}\frac{dx(\omega)}{dt} &= s(\omega)(y(\omega) - x(\omega)) \\ \frac{dy(\omega)}{dt} &= r(\omega)x(\omega) - y(\omega) - x(\omega)z(\omega) \\ \frac{dz(\omega)}{dt} &= x(\omega)y(\omega) - b(\omega)z(\omega),\end{aligned}\tag{8}$$

which need to be integrated in time to obtain the evolution of the stochastic state vector  $u(\omega) = (x(\omega), y(\omega), z(\omega))^T$ .

For sampling approaches to UQ and filtering—such as the EnKF—the parameters are sampled according to Eq. (7). Each sample can be used to integrate forward in time some initial conditions  $u_0$  using Eq. (6) and the—for this model common—fourth order Runge-Kutta scheme (Lorenz, 1963).

For the PCE-based approach discussed in this work we directly use the truncated PCE as representation for the involved RVs—an approach also followed in Shen

et al. (2010); Pajonk et al. (2011) for the Lorenz-84 model. We replace the parameters  $s(\omega)$ ,  $r(\omega)$  and  $b(\omega)$  by the truncated Hermite transforms (see A)  $\widehat{\sigma}$ ,  $\widehat{\rho}$ , and  $\widehat{\beta}$ , where  $\widehat{\cdot}$  denotes the projection on the finite subspace generated by  $\{H_\alpha | \alpha \in \mathcal{J}_Z\}$  and  $\sigma := \mathcal{H}(s)$ ,  $\rho := \mathcal{H}(r)$ ,  $\beta := \mathcal{H}(b)$ . The state variables  $x, y$ , and  $z$  are treated similarly; their transforms are denoted as  $\xi := \mathcal{H}(x)$ ,  $\eta := \mathcal{H}(y)$ , and  $\zeta := \mathcal{H}(z)$ . For simplicity the same subspace is used for all parameters and variables. Due to the truncation the equations Eq. (8) cannot be satisfied exactly anymore: for example the result of a product of two truncated PCEs  $Q_2(\widehat{\xi}, \widehat{\eta})$  (see A) does not necessarily lie in that subspace anymore. To solve this problem we perform a Galerkin projection back onto that subspace. The final result is then the stochastic evolution equation Eq. (8) projected onto the subspace:

$$\begin{aligned} \frac{d\widehat{\xi}}{dt} &= \widehat{Q}_2(\widehat{\sigma}, \widehat{\eta} - \widehat{\xi}) \\ \frac{d\widehat{\eta}}{dt} &= \widehat{Q}_2(\widehat{\rho}, \widehat{\xi}) - \widehat{\eta} - \widehat{Q}_2(\widehat{\xi}, \widehat{\xi}) \\ \frac{d\widehat{\zeta}}{dt} &= \widehat{Q}_2(\widehat{\xi}, \widehat{\eta}) - \widehat{Q}_2(\widehat{\beta}, \widehat{\xi}), \end{aligned} \quad (9)$$

This system of SDEs can be integrated forward in time using the same Runge-Kutta scheme already mentioned.

### 3.2 Experimental Setup

As initial condition we choose

$$u_0 := (1.508870, -1.531271, 25.46091)^T \quad (10)$$

(Miller et al., 1994; Evensen, 2009a) for the “truth” state for all experiments. Evidence is simulated by taking noisy measurements of  $u$  of this “truth” case with the classical choice of parameters (Lorenz, 1963):

$$s_{truth} := 10, \quad r_{truth} := 28, \quad b_{truth} := 8/3. \quad (11)$$

The first guess of parameters for the estimation is chosen as

$$\begin{aligned} s(\omega) &\sim \mathcal{N}(9.6, 0.48), \\ r(\omega) &\sim \mathcal{N}(28.5, 1.425), \\ b(\omega) &\sim \mathcal{N}(2.55, 0.1275). \end{aligned} \quad (12)$$

and the initial state is—assuming lack of better knowledge—given by a standard normal distribution for each component

$$x_0(\omega), y_0(\omega), z_0(\omega) \sim \mathcal{N}(0, 1). \quad (13)$$

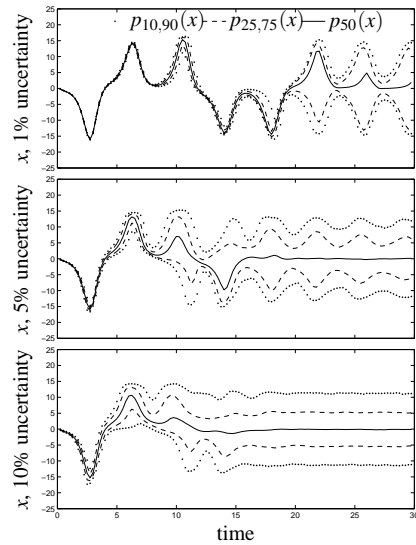


Figure 1: Growth of different initial parametric uncertainties (1%, 5%, 10%) in  $s(\omega)$ ,  $r(\omega)$  and  $b(\omega)$ , starting from a practically negligible initial uncertainty in the state of  $x_0(\omega), y_0(\omega), z_0(\omega) \sim \mathcal{N}(0, 10e - 16)$ . Some quantiles of the pdf of the  $x$ -component are plotted over time. The quantiles are estimated from 10000 random samples. The plot is similar for the  $y$  and  $z$  components.

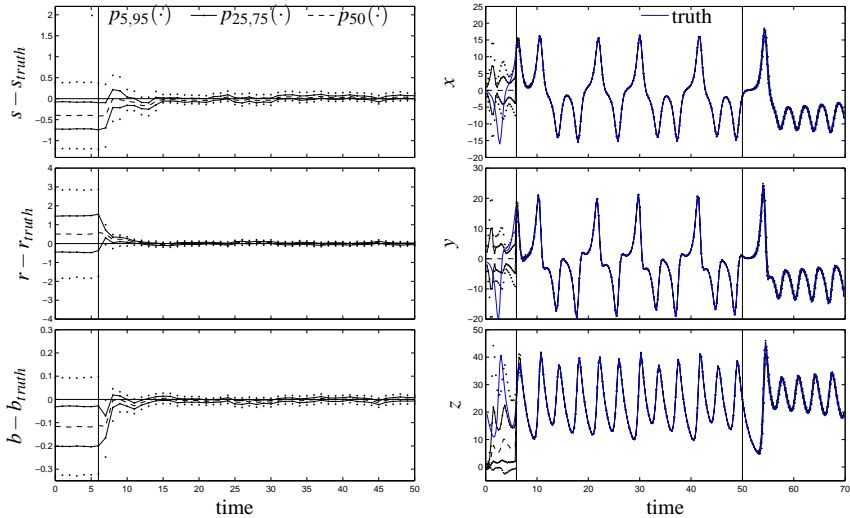


Figure 2: Evolution of the parameter and state estimates for the PCE-based method with polynomial order 2. The vertical bar marks the first update, the second marks the last update.

The mean parameter values clearly differ from the truth Eq. (11), and we assume a 5% initial standard deviation for each parameter. Note that this first guess distribution “covers” the “truth” parameters. However, due to the chaotic nature of the Lorenz-63 model, these relatively small errors quickly cause major uncertainty on the state and a significant diversion from the “truth” (this is demonstrated in Fig. 1). Besides, the significant uncertainty in the first guess state—Eq. (13)—will pose a significant challenge to the estimation algorithms.

No updates occur for the first 6 time units of all assimilation experiments to allow the parametric uncertainty to mix into the state. Then, noisy measurements of the state are taken and assimilated once every time unit until time unit 50, where a 20 forecast period of time units starts. The measurement noise is  $\varepsilon \sim \mathcal{N}(0, 0.1)$  with i.i.d. samples for each state vector entry, unless stated otherwise. The measurement noise samples are *exactly* the same for all experiments to allow for direct comparison. The state as well as the parameters are updated—but let us stress once more that only the state is measured. All numerical integration is performed using a fourth-order Runge-Kutta scheme with a time-stepping of  $\delta t = 0.25$  dimensionless time units.

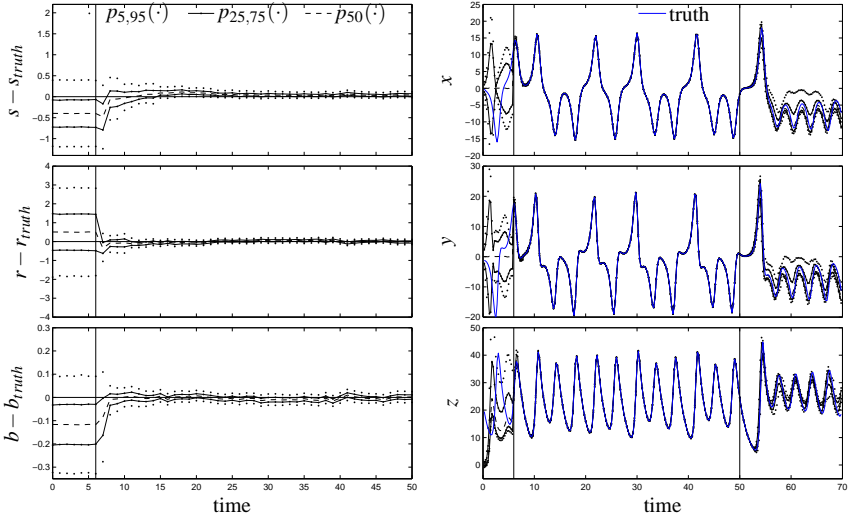


Figure 3: Evolution of the parameter and state estimates for the PCE-based method with polynomial order 3. The vertical bar marks the first update, the second marks the last update.

### 3.3 Results and Discussion

The previously described identification experiment has been conducted with both the PCE-based identification approach and a straightforward implementation of the so-called “perturbed observation” EnKF for comparison Evensen (2009a,b). Figs. 2–4 show some representative evolutions of the parameter and state identification process, given by some percentile estimates obtained from the PCE or the ensemble, respectively. The computational time used to perform the EnKF run with 50 ensemble members (around 7.5 seconds on a year 2008 laptop) is approximately the same as for the PCE-based method with order two (around 7.6 seconds), which is why this ensemble size has been chosen for visual comparison. Fig. 5 aims at summarising the final parameter identification results by presenting some probability density estimates obtained at the final parameter update,  $t = 50$ . The estimates have been obtained by a kernel density technique (Botev et al., 2010).

From Figs. 2 and 3 it may be seen that the PCE-based method generally retains a larger variance when compared to the EnKF, especially for the important parameter  $r$ . This is supported by Fig. 5. For the EnKF, the variance generally drops much faster. This is also true across different ensemble sizes, up to the size of 10000, therefore additional evolution plots have been omitted. For ensemble sizes of, say,

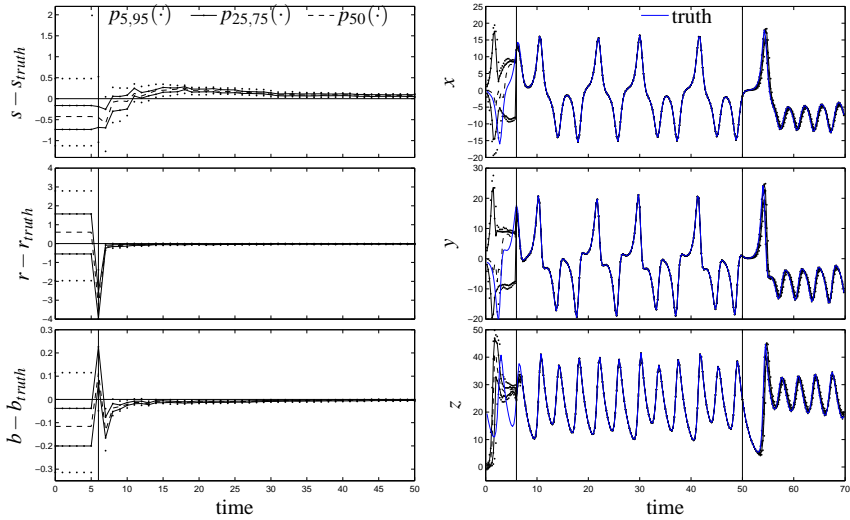


Figure 4: Evolution of the parameter and state estimates for the EnKF with ensemble size 50. The vertical bar marks the first update, the second marks the last update.

below 50, the performance of the EnKF strongly depends on the initial ensemble: all other things being the same, for quite some initial ensembles the EnKF even may diverge. A—less severe—example can be seen in Fig. 5 for the  $N = 30$  case. This dependence on the initial sample naturally does not exist for the PCE-based method, since it is sampling-free. On the other hand, the PCE-based method does not seem to strongly stabilise over the estimation period: there are always minor updates to the mean values. However, the variance stabilises after some time.

Another thing worth mentioning is that for the PCE-based method, in the beginning of the estimation period at  $t = 6$ , no immediate strong corrections to the parameters occur like with the EnKF (the effect also occurs for larger ensemble sizes like  $N = 10000$ , so Fig. 4 is representative in that sense). The PCE-based method first corrects the—severely wrong—state estimates for two to three updates and only then starts to correct the parameters (see  $t \in [6, 10]$  of Fig. 2). This is especially true for the lesser important parameters,  $s$  and  $b$ . On the other hand, the EnKF moves the parameter and state estimates quite far out of the reliability region of the estimator: e.g. in Fig. 4, where  $r_{truth}$  is suddenly in a very low probability region. Also the state estimates may be quite off in the initial phase (see Fig. 4). While this is corrected by the next updates, in a sense the the PCE-based method seems to be more reliable: the uncertainty in the estimator never severely underestimates the actual error. This may hint at advantages of the PCE-based method for non-linear



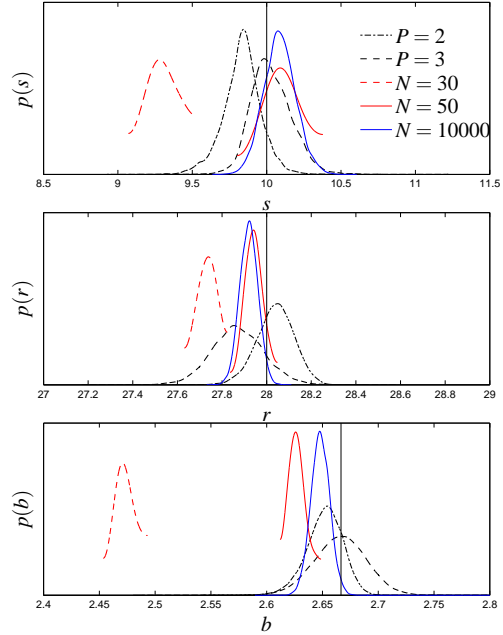


Figure 5: Probability density estimates for  $t = 50$  (after the last update), for different methods:  $P$  denotes the polynomial order of the PCE-based approach, whereas  $N$  denotes the EnKF results. The to-be-identified correct values  $s_{truth}$ ,  $r_{truth}$  and  $b_{truth}$  are marked with a vertical bar.

applications, where usually iterative EnKF approaches are employed (e.g. [Annan & Hargreaves \(2004\)](#)).

From the results of the 10000-member EnKF run in Fig. 5 one can assume that this experiment generally is not a simple task for linear approximations to the conditional expectation. Even for this quite large ensemble, the final estimates are not centred on the true value. This is likely due to the severe non-linearity of the forward model that transports information in the form of uncertainty from the parameters to the states—and these states are the only source of information available to the estimation procedure.

## 4 Conclusion

In this paper a fully deterministic implementation of the linear approximation to the conditional expectation has been presented. It is based on spectral series expansions—like Wiener’s polynomial chaos expansion—of the random variables that describe uncertainties in a system of interest. This estimation procedure has been applied to a low-dimensional combined parameter and state estimation problem of a chaotic system, with the focus set on parameter estimation. There, it has been compared to a close relative, the ensemble Kalman filter.

The application was successful in the sense that improved parameter estimates—together with sensible variance estimates representing the residual uncertainty—could be derived, even for this non-linear problem. Such estimates could already be obtained by a relatively modest order of polynomial expansion of  $P = 2$ , and could be improved with  $P = 3$ . For these settings, the PCE-based method seems to be slightly more reliable in terms of variance estimation when compared to EnKF with similar computational load.

The larger variance of the estimates—compared to EnKF results—likely has a two-fold reason. One is certainly the well-known underestimation of variance by the EnKF due to sampling errors. However, another reason is likely a result of a simplifying implementation assumption in the PCE-based method. Currently, the additional information introduced by the evidence lives in the same space as the simulated data. This may result in a smaller reduction of variance and is currently being investigated.

## A The Hermite Algebra

Consider first the usual univariate Hermite polynomials  $\{h_k\}$  as defined in [Hida et al. \(1999\)](#); [Holden et al. \(2010\)](#); [Janson \(1997\)](#); [Malliavin \(1997\)](#). As the univariate Hermite polynomials are a linear basis for the polynomial algebra, i.e. every

polynomial can be written as linear combination of Hermite polynomials, this is also the case for the product of two Hermite polynomials  $h_k h_\ell$ , which is clearly also a polynomial:

$$h_k(t)h_\ell(t) = \sum_{n=|k-\ell|}^{k+\ell} c_{k\ell}^{(n)} h_n(t) \quad (14)$$

The coefficients are only non-zero for integer  $g = (k + \ell + n)/2 \in \mathbb{N}$  and if  $g \geq k \wedge g \geq \ell \wedge g \geq n$  (Malliavin, 1997). They can be explicitly given

$$c_{k\ell}^{(n)} = \frac{k! \ell!}{(g-k)!(g-\ell)!(g-n)!}, \quad (15)$$

and are called the structure constants of the univariate Hermite algebra.

For the multivariate Hermite algebra, analogous statements hold (Malliavin, 1997):

$$H_\alpha(\mathbf{t})H_\beta(\mathbf{t}) = \sum_{\gamma} c_{\alpha\beta}^{\gamma} H_{\gamma}(\mathbf{t}). \quad (16)$$

with the multivariate structure constants

$$c_{\alpha\beta}^{\gamma} = \prod_{j=1}^{\infty} c_{\alpha_j\beta_j}^{\gamma_j}, \quad (17)$$

defined in terms of the univariate structure constants Eq. (15).

With these structure constants one defines the matrices  $Q_2^{\gamma} := (c_{\alpha\beta}^{\gamma})$  with multi-indices  $\alpha$  and  $\beta$ . With this notation the Hermite transform of the product of two random variables  $r_1(\omega) = \sum_{\alpha \in \mathcal{J}} r_1^{\alpha} H_{\alpha}(\theta)$  and  $r_2(\omega) = \sum_{\beta \in \mathcal{J}} r_2^{\beta} H_{\beta}(\theta)$  is

$$\mathcal{H}(r_1 r_2) = ((r_1) Q_2^{\gamma} (r_2)^T)_{\gamma \in \mathcal{J}}. \quad (18)$$

Each coefficient is a bilinear form in the coefficient sequences of the factors, and the collection of all those bilinear forms  $Q_2 = (Q_2^{\gamma})_{\gamma \in \mathcal{J}}$  is a bilinear mapping that maps the coefficient sequences of  $r_1$  and  $r_2$  into the coefficient sequence of the product

$$\begin{aligned} \mathcal{H}(r_1 r_2) &= Q_2((r_1), (r_2)) \\ &= Q_2(\mathcal{H}(r_1), \mathcal{H}(r_2)). \end{aligned} \quad (19)$$

Products of more than two random variables may now be defined recursively through the use of associativity. e.g.  $r_1 r_2 r_3 r_4 = (((r_1 r_2) r_3) r_4)$ :

$$\begin{aligned} \forall k > 2: \quad \mathcal{H}\left(\prod_{j=1}^k r_j\right) &:= Q_k((r_1), (r_2), \dots, (r_k)) := \\ &Q_{k-1}(Q_2((r_1), (r_2)), (r_3), \dots, (r_k)). \end{aligned} \quad (20)$$

Each  $Q_k$  is again composed of a sequence of  $k$ -linear forms  $\{Q_k^\gamma\}_{\gamma \in \mathcal{J}}$ , which define each coefficient of the Hermite transform of the  $k$ -fold product.

## B Higher Order Moments

Consider RVs  $\mathbf{r}_j(\omega) = \sum_{\alpha \in \mathcal{J}} \mathbf{r}_j^\alpha H_\alpha(\theta(\omega))$  with values in a vector space  $\mathcal{V}$ , then  $\bar{\mathbf{r}}_j$ ,  $\tilde{\mathbf{r}}_j(\omega)$ , as well as  $\mathbf{r}_j^\alpha$  are in  $\mathcal{V}$ . Any moment may be easily computed knowing the PCE. The  $k$ -th centred moment is defined as

$$\mathbf{M}_{\mathbf{r}_1 \dots \mathbf{r}_k}^k = \mathbb{E} \left( \bigotimes_{j=1}^k \tilde{\mathbf{r}}_j \right), \quad (21)$$

a tensor of order  $k$ . Thus it may be expressed via the PCE as

$$\mathbf{M}_{\mathbf{r}_1 \dots \mathbf{r}_k}^k = \sum_{\gamma^1, \dots, \gamma^k \neq 0} \mathbb{E} \left( \prod_{j=1}^k H_{\gamma^j}(\theta) \right) \bigotimes_{m=1}^k \mathbf{r}_m^{\gamma^m}, \quad (22)$$

and in particular:

$$\begin{aligned} \mathbf{C}_{\mathbf{r}_1 \mathbf{r}_2} &= \mathbf{M}_{\mathbf{r}_1 \mathbf{r}_2}^2 = \mathbb{E}(\mathbf{r}_1 \otimes \mathbf{r}_2) \\ &= \sum_{\gamma, \beta > 0} \mathbb{E}(H_\gamma H_\beta) \mathbf{r}_1^\gamma \otimes \mathbf{r}_2^\beta \\ &= \sum_{\gamma > 0} \gamma! \mathbf{r}_1^\gamma \otimes \mathbf{r}_2^\gamma, \end{aligned} \quad (23)$$

as  $\mathbb{E}(H_\gamma H_\beta) = \delta_{\gamma\beta} \gamma!$ . The expected values of the products of Hermite polynomials in Eq. (22) may be computed analytically, by using the formulas from A.

## References

- Ambadan, J. T. & Tang, Y. (2009). Sigma-point kalman filter data assimilation methods for strongly nonlinear systems. *Journal of the Atmospheric Sciences*, 66(2), 261–285.
- Andrews, A. (1968). A square root formulation of the Kalman covariance equations. *AIAA Journal*, 6(6), 1165–1166.
- Annan, J. D. & Hargreaves, J. C. (2004). Efficient parameter estimation for a highly chaotic system. *Tellus A*, 56(5), 520–526.
- Blanchard, E. D., Sandu, A., & Sandu, C. (2010). A polynomial chaos-based kalman filter approach for parameter estimation of mechanical systems. *Journal of Dynamic Systems, Measurement, and Control*, 132(6), 061404.

- Bobrowski, A. (2005). *Functional Analysis for Probability and Stochastic Processes*. Cambridge, UK: Cambridge University Press.
- Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel density estimation via diffusion. *Annals of Statistics*, 38(5), 2916–2957.
- El-Moselhy, T. A. (2010). *Field Solver Technologies for Variation-Aware Interconnect Parasitic Extraction*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.
- Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Evensen, G. (2009a). *Data Assimilation — The Ensemble Kalman Filter* (second ed.). Berlin, Germany: Springer-Verlag.
- Evensen, G. (2009b). The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Systems Magazine*, 29, 82–104.
- Fleming, R. J. (1993). The dynamics of uncertainty: application to parameterization constants in climate models. *Climate Dynamics*, 8(3), 135–150.
- Goldstein, M. & Wooff, D. (2007). *Bayesian Linear Statistics—Theory and Methods*. Chichester: John Wiley & Sons.
- Hamill, T. M., Whitaker, J. S., Anderson, J. L., & Snyder, C. (2009). Comments on "sigma-point Kalman filter data assimilation methods for strongly nonlinear systems". *Journal of the Atmospheric Sciences*, 66(11), 3498–3500.
- Hida, T., Kuo, H.-H., Potthoff, J., & Streit, L. (1999). *White Noise—An Infinite Dimensional Calculus*. Kluwer, Dordrecht.
- Hida, T. & Si, S. (2008). *Lectures on white noise functionals*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Holden, H., Øksendal, B., Ubøe, J., & Zhang, T. (2010). *Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach* (second ed.). Universitext. Berlin, Germany: Springer-Verlag.
- Janson, S. (1997). *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge, UK: Cambridge University Press.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transaction of the ASME - Journal of Basic Engineering*, 82(Series D), 35–45.

- Kivman, G. A. (2003). Sequential parameter estimation for stochastic systems. *Nonlinear Processes in Geophysics*, 10(3), 253–259.
- Kolmogorov, A. N. (1956). *Foundations of the Theory of Probability* (second english ed.). New York: Chelsea Publishing Company.
- Krosche, M. & Niekamp, R. (2010). Low rank approximation in spectral stochastic finite element method with solution space adaption. Technical Report 2010-3, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig, Germany.
- Lahoz, W., Khattatov, B., & Ménard, R. (Eds.). (2010). *Data Assimilation: Making Sense of Observations*. Berlin: Springer-Verlag.
- Le Maître, O. P. & Knio, O. M. (2010). *Spectral Methods for Uncertainty Quantification with Applications to Computational Fluid Dynamics*. Scientific Computation. Springer.
- Lea, D. J., Allen, M. R., & Haine, T. W. N. (2000). Sensitivity analysis of the climate of a chaotic system. *Tellus A*, 52(5), 523–532.
- Li, J. & Xiu, D. (2009). A generalized polynomial chaos based ensemble Kalman filter with high accuracy. *Journal of Computational Physics*, 228(15), 5454–5469.
- Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2), 130–141.
- Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. Chichester: John Wiley & Sons.
- Malliavin, P. (1997). *Stochastic Analysis*. Springer: Springer-Verlag, Berlin.
- Matthies, H. G. (2007). *Encyclopedia of Computational Mechanics*, chapter Uncertainty Quantification with Stochastic Finite Elements. Chichester: John Wiley & Sons.
- Matthies, H. G., Litvinenko, A., Pajonk, O., Rosić, B. V., & Zander, E. (2011). Parametric and uncertainty computations with tensor product representations. In Dienstfrey, A. & Boisvert, R. (Eds.), *Uncertainty Quantification in Scientific Computing*, IFIP Advances in Information and Communication Technology, Berlin. Springer.
- Miller, R. N., Ghil, M., & Gauthiez, F. (1994). Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences*, 51(8), 1037–1056.

- Moolenaar, H. E. & Selten, F. M. (2004). Finding the effective parameter perturbations in atmospheric models: the Lorenz63 model as case study. *Tellus A*, 56(1), 47–55.
- Nouy, A. & Le Maître, O. P. (2009). Generalized spectral decomposition for stochastic nonlinear problems. *Journal of Computational Physics*, 228(1), 202–235.
- Øksendal, B. K. (2003). *Stochastic Differential Equations—An Introduction with Applications* (sixth ed.). Berlin, Germany: Springer-Verlag.
- Pajonk, O., Rosić, B. V., Litvinenko, A., & Matthies, H. G. (2011). A deterministic filter for non-Gaussian Bayesian estimation. Informatikbericht 2011-04, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig.
- Pence, B., Fathy, H., & Stein, J. (2010). A maximum likelihood approach to recursive polynomial chaos parameter estimation. In *American Control Conference (ACC)*, (pp. 2144–2151).
- Rosić, B. V., Litvinenko, A., Pajonk, O., & Matthies, H. G. (2011). Direct Bayesian update of polynomial chaos representations. Informatikbericht 2011-02, Institut für Wissenschaftliches Rechnen, Technische Universität Braunschweig.
- Saad, G. & Ghanem, R. (2009). Characterization of reservoir simulation models using a polynomial chaos-based ensemble Kalman filter. *Water Resources Research*, 45(W04417), 1–19.
- Sakov, P. & Oke, P. R. (2008). A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A*, 60(2), 361–371.
- Shen, C. Y., Evans, T. E., & Finette, S. (2010). Polynomial chaos quantification of the growth of uncertainty investigated with a Lorenz model. *Journal of Atmospheric and Oceanic Technology*, 27(6), 1059–1071.
- Tang, Y. & Ambadan, J. (2009). Reply. *Journal of the Atmospheric Sciences*, 66(11), 3501–3503.
- Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, USA: Society for Industrial and Applied Mathematics.
- Tippett, M. K., Anderson, J. L., & Bishop, C. H. (2003). Ensemble square root filters. *Monthly Weather Review*, 131, 1485 – 1490.
- Whittle, P. (2000). *Probability via expectation* (fourth ed.). Berlin, Germany: Springer-Verlag.

Wiener, N. (1938). The homogeneous chaos. *American Journal of Mathematics*, 60(4), 897–936.

Xiu, D. (2010). *Numerical Methods for Stochastic Computations—A Spectral Method Approach*. Princeton, NJ: Princeton University Press.



## Informatikberichte

2011-04	O. Pajonk, B.V. Rosić, A. Litvinenko, and H. G. Matthies	A Deterministic Filter for non-Gaussian Bayesian Estimation
2011-05	H. G. Matthies	A Hitchhiker's Guide to Mathematical Notation and Definitions
2011-06	R. van Glabbeek, U. Goltz, J.-W. Schicke	On Causal Semantics of Petri Nets
2011-07	H. Cichos, S. Oster, M. Lochau, A. Schürr	Extended Version of Model-based Coverage-Driven Test Suite Generation for Software Product Lines
2011-08	W.-B. Pöttner, J. Morgenroth, S. Schildt, L. Wolf	An Empirical Performance Comparison of DTN Bundle Protocol Implementations
2011-09	H. G. Matthies, A. Litvinenko, O. Pajonk, B. V. Rosić and E. Zander	Parametric and Uncertainty Computations with Tensor Product Representations
2011-10	B. V. Rosić, A. Kučerová, J. Sýkora, A. Litvinenko, O. Pajonk and H. G. Matthies	Parameter Identification in a Probabilistic Setting
2011-11	M. Espig, W. Hackbusch, A. Litvinenko, H. G. Matthies and E. Zander	Efficient Analysis of High Dimensional Data in Tensor Formats
2011-12	S. Oster	A Semantic Preserving Feature Model to CSP Transformation
2012-01	O. Pajonk, B. V. Rosić and H. G. Matthies	Deterministic Linear Bayesian Updating of State and Model Parameters for a Chaotic Model